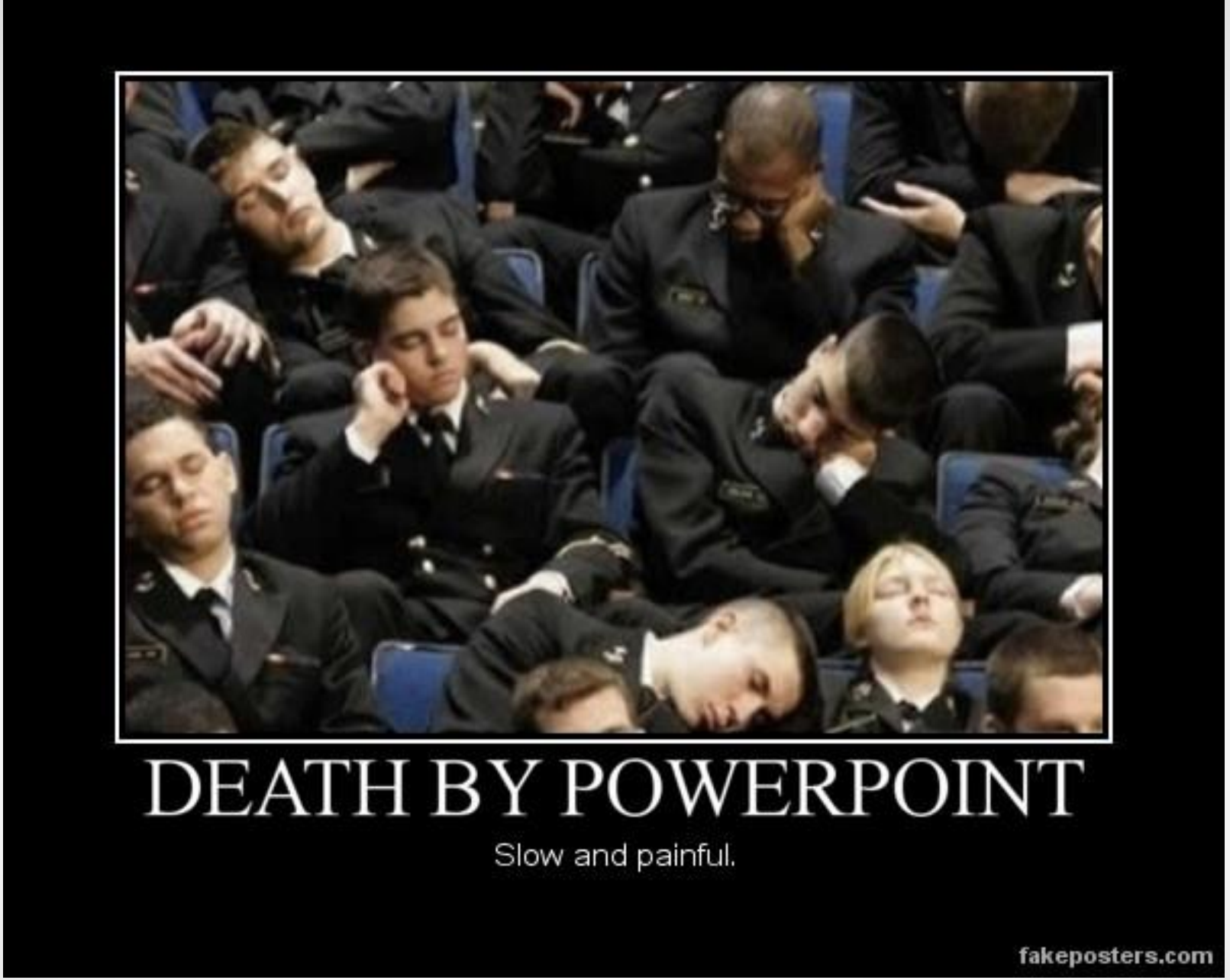
GHRC User Working Group Meeting WELCOME

September 25-26, 2014

Huntsville, AL



Source: http://www.fakeposters.com/posters/death-powerpoint/

Agenda – Day 1

0800 - 0830	*Coffee; Meet and Greet*	
0830 – 0845	GHRC Welcome UAH welcome	Ramachandran Graves
0845 - 0945	Opening Comments • Meeting Overview & Logistics • View from HQ • View from ESDIS	 Harrison Kakar Berrick
0945 – 1030	GHRC DAAC Overview Presentation • Management & Organization • Vision DISCUSSION	 Conover Ramachandran
1030 - 1100	*Break - SPoRT Tour*	Zavodsky
1100 - 1200	Infrastructure • Hardware and Processing Automation • Data Search, Access and Order Tools DISCUSSION • Web presence DISCUSSION	 Ellett, Beaumont Harrison Smith, Flynn
1200 - 1300	*Working Lunch*	
1300 - 1430	Data and Related Tools • Satellite and Airborne Microwave Sensors ○ Data overview ○ Regional Air Sea Interactions Tool DISCUSSION • Field Campaigns ○ Data overview ▪ Hurricane Science ▪ GPM Ground Validation ○ Mission Coordination Portal ○ HS3 Data System DISCUSSION	 Drewry Keiser Bugbee Hawkins He Maskey
1430 - 1500	*Break*	

Agenda – Day 1 (cont'd)

1500 - 1700	UWG Member presentations (10 minutes each) Microwave Research • Assessing Global Water and Energy Budgets • Merging retrievals for passive microwave imagers and sounders Lightning Research • Lightning research at MSFC • Monitoring the performance of space-based optical lightning sensors • Multi-sensor data workflows for lightning science Hurricane Science • Global distributions of various types of precipitation systems from radar and passive microwave observations • The Cyclone Global Navigation System (CYGNSS) Earth Venture Mission • Contributions of GHRC datasets to the Zipser Tropical Convection Research Group GPM Ground Validation • GPM Field Campaigns Applications • Integration of Earth Science Research and Education at UAH • SPoRT/MSFC Applied Science: Disaster Activities	 Kummerow Duncan Blakeslee Buechler Bruning Liu Ruf Zawislak Wolff Griffin Molthan
1700-1730	Summary, Actions and Wrap-up	Ramachandran
1730	*Adjourn*	
1830	*Group Dinner - 1892 East*	

Agenda – Day 2

0800 - 0830	*Coffee; Meet and Greet*	
0830 - 0845	Welcome, recap, plan for day	Ramachandran
0845 - 0930	Lightning Data and Tools • Data overview • LIS Space Time Search • GLM Validation Tool DISCUSSION	Regner
0930 -1000	AMSR SIPS highlights • LANCE AMSR2 • AMSR-E Provenance Browser DISCUSSION	Lin McEniry
1000 - 1020	ESDIS initiatives DISCUSSION	Conover
1020 - 1030	Charge to UWG	Ramachandran
1030 - 1100	*Break - SERVIR Tour*	Irwin
1100 - 1330	Executive Session - discussion and report *(working lunch provided)*	UWG
1330 – 1430	Board Advice to DAAC - prioritization of current work and future efforts	All
1430 – 1500	Summary of Actions and Meeting Wrap-up	Ramachandran
1500	*Adjourn*	

GLOBAL HYDROLOGY RESOURCE CENTER

A NASA Distributed Active Archive Center

Rahul Ramachandran
DAAC Manager
rahul.ramachandran@nasa.gov

Helen Conover
GHRC Operations Manager
hconover@itsc.uah.edu

Presented at the GHRC User Working Group Meeting
September 25-26, 2014

NASA

GHRC Global Hydrology Resource Center

itsc UAH THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Global Hydrology Resource Center

- **Full service data center** providing data ingest, routine and custom processing, archive, distribution, user support, and science data services
- **Collaboration** between NASA and the University of Alabama in Huntsville to **infuse advanced information technologies** to a variety of **science data projects**
- Global **lightning data from space**, airborne and ground based observations from **hurricane science field campaigns** and Global Precipitation Mission **(GPM) ground validation** experiments, and **satellite passive and active microwave** products

http://ghrc.nsstc.nasa.gov/

What we do
Data Stewardship

CURATION
Work with Science Teams to gather not only data but also **all relevant information**

DOCUMENTATION
Capture this information to **create a knowledge base** for our stake holder communities

PRESERVATION
Follow **documented policies** and **engineered** procedures at every step to insure information preservation against all reasonable contingencies

INTEROPERABILITY STANDARDS
Ensure that the information is "**independently understandable**" to all stakeholders without requiring experts

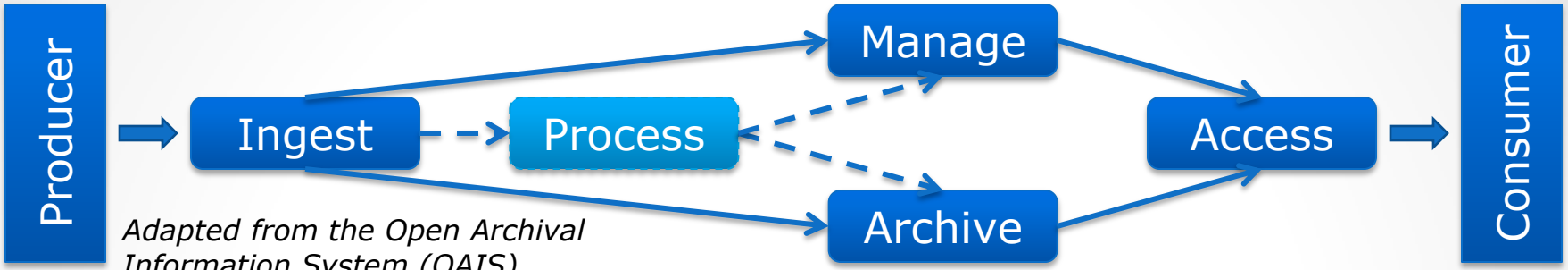
PROCESSING
include **science product generation** and reformatting, algorithm integration and test, interfaces with external providers

NASA's Earth science data stewards for scientific, educational, commercial and governmental communities, with a focus on data for the global hydrologic cycle

History

- **Marshall DAAC** was established in 1991 at the beginning of NASA's EOSDIS program
  - Based on the WetNet project led by Michael Goodman and a local science data management effort led by Sara Graves.
  - Science focus was passive microwave and lightning data
- **LIS Enhanced Science Computing Facility (E-SCF)** was established 1997 to manage data from the Lightning Imaging Sensor on TRMM
  - Co-branded as **Global Hydrology Resource Center**
  - Funding through the MSFC lightning science team
  - Supplemental funding through other science projects (e.g., the Hurricane Science Program for specific field campaigns)
- **GHRC DAAC** was added to the NASA Earth Science Data and Information Systems (ESDIS) project for core funding in 2009
- **AMSR-E SIPS at GHRC** was established in 1998 to generate standard products from the AMSR-E instrument on Aqua. Near-real time processing for **LANCE** was added in 2010.

Data Center Operations



Adapted from the Open Archival Information System (OAIS) model

- **Ingest and archive** data and metadata, orbit/altitude data, documents, algorithms, instrument and spacecraft history, ancillary data from external sources for production.
- **Processing** including science product generation and reformatting, algorithm integration and test, interfaces with external providers (EOSDIS, other data centers).
- **Data discovery and access services** include direct online access to most data products, an online search and order system, registration of all data in NASA Earth science data catalogs, and support for a variety of data access web services
- **Data distribution and user services** including processing orders (subscriptions and on-demand), tracking orders across system, prioritizing based on resource management policy.

GHRC Staffing Profile

Except for DAAC Manager, all GHRC staff are matrixed from UAH's Information Technology and Systems Center.

Mission and Science Support

- Metadata Development
- Documentation
- Science Team Interactions

User Services

- Customer Interactions
- Web Site and Social Media

Operations

- Ingest, Processing and Archive Management
- Systems Testing

Engineering

- Systems Engineering
- Software Engineering and Development
- HS3 Data System Planning
- Evolution of Existing Systems

Infrastructure

- Systems Administration
- Database Administration

IT Security

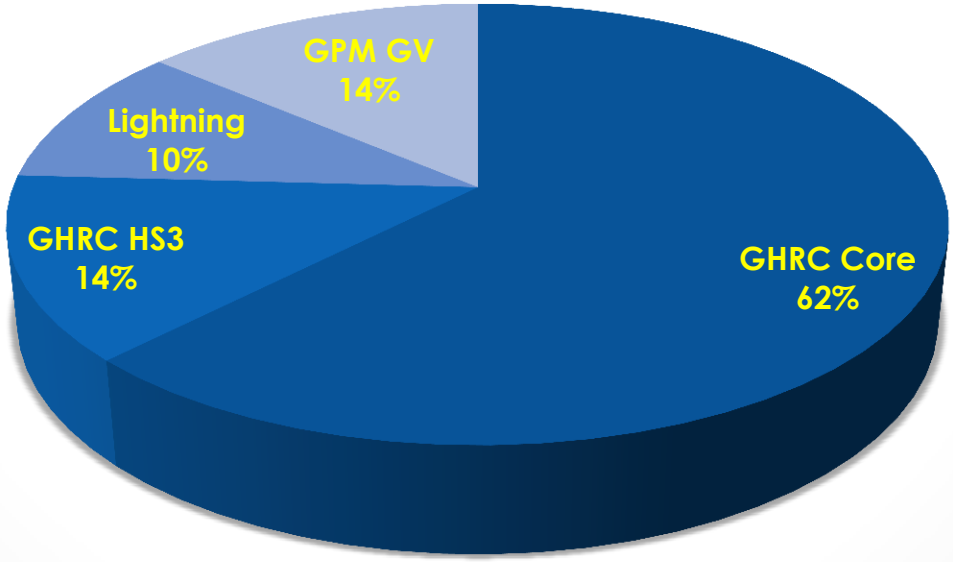
Project Management

~6 WYE spread over ~20 people

GHRC Global Hydrology Resource Center

Shared Resources

- GHRC leverages supplemental funding from science projects to provide data management services using GHRC infrastructure
- GHRC also shares staff with the AMSR SIPS



Combined funding snapshot for 2014

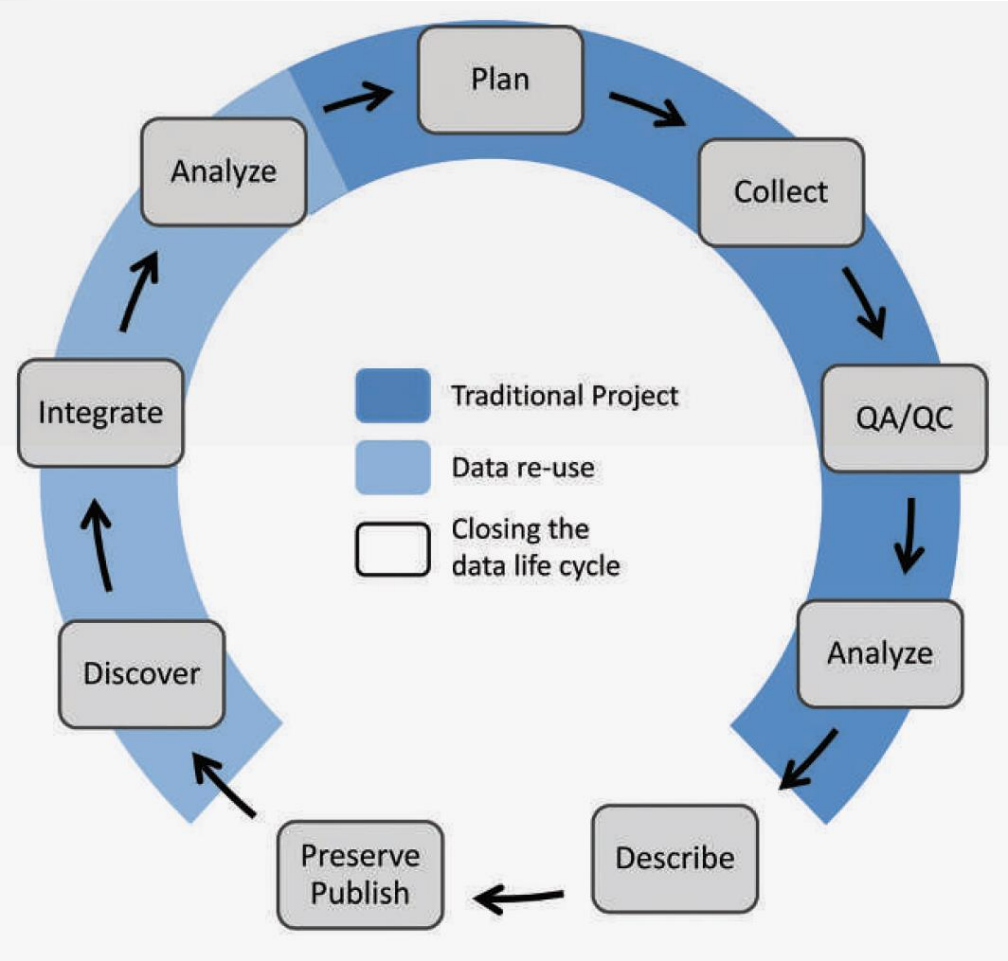
Importance of (Open) Data

- Fair Access to Science and Technology Research Act (FASTR) introduced in both the Senate and the House in 2013.
- OSTP memorandum :

  "directs each Federal agency with over $100 million in annual conduct of research and development expenditures to develop a plan to support increased public access to the results of research funded by the Federal Government. This includes any results published in peer-reviewed scholarly publications that are based on research that directly arises from Federal funds . . ." (OSTP 2013, 2).
- Explicitly states that "such results include peer-reviewed publications and digital data."

Asher, A., K. Deards, M.Esteva, M. Halbert, L. Jahnke, C. Jordan, S. D.C. Keralis, et al. 2013. "Research Data Management: Principle, Practices, and Prospects". Washington DC, USA.

Role of DAAC's in the Data Life Cycle



- Provides two steps needed to complete the data lifecycle
- Enables data to retain value past the life of the project and creates new research/application opportunities

Figure Source: Ruegg, J., C. Gries, B. Bond-Lamberty, G. J. Brown, B. S. Felzer, N. E. McIntyre, P. A. Soranno, K. L. Vanderbilt, and K. C. Weathers. 2014. "Completing the Data Life Cycle: Using Information Management in Macrosystems Ecology Research." Frontiers in Ecology and the Environment 12 (1): 24–30. doi:10.1890/210375.

GHRC Mission Statement

- To serve as NASA's Earth science *data stewards* for scientific, educational, commercial and governmental communities, with a focus on data for the global hydrologic cycle
  - Hydrologic Cycle
  - Severe Weather Interactions
  - Lightning
  - Atmospheric Convection
- To provide *knowledge augmentation services* encompassing tools, infrastructure, user support, and expertise to our stakeholders

What we do
Knowledge Augmentation Services

PROVENANCE
Make the preserved data/information available to all our stakeholder communities with traceability to support authenticity (***AMSR-E Provenance***)

DATA DISCOVERY
Develop new tools for data discovery, curation and aggregation (***LIS Interactive Browse***)

FIELD CAMPAIGN INFRASTRUCTURE
Create specialized portals for managing field campaigns and collecting data (***Field Campaign Portal***)

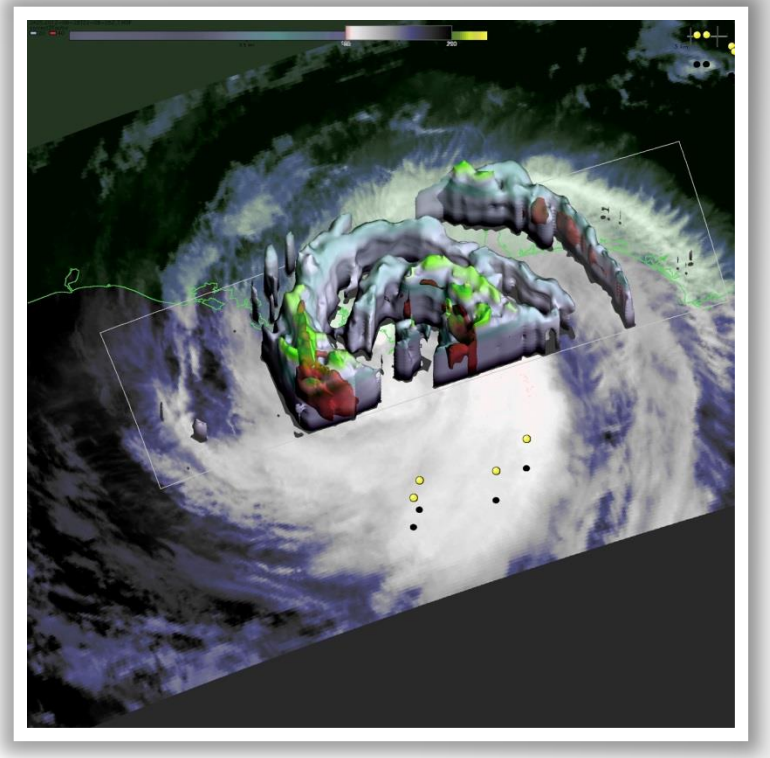
DATA USE
Develop new tools for access, analysis and visualization (***HS3 Data System, GLM Validation Tool, RASI***)

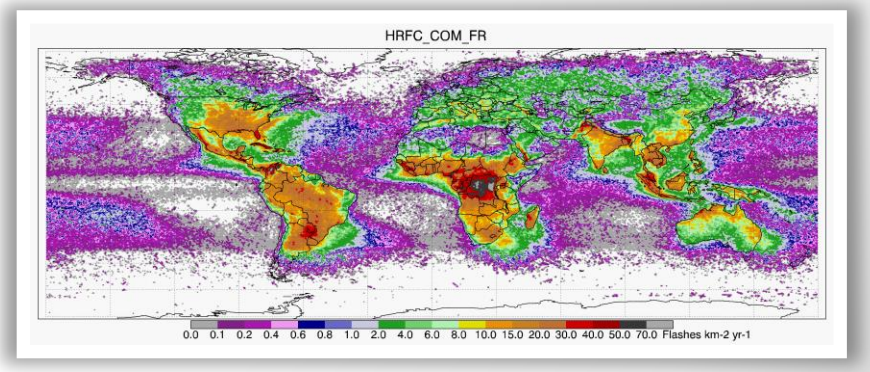
INFUSING CUTTING EDGE INFORMATICS
Research new approaches and technologies and infuse them into operational processes

GHRC provides knowledge augmentation services encompassing tools, infrastructure, user support, and expertise to our stakeholders

What we serve
Lightning Data

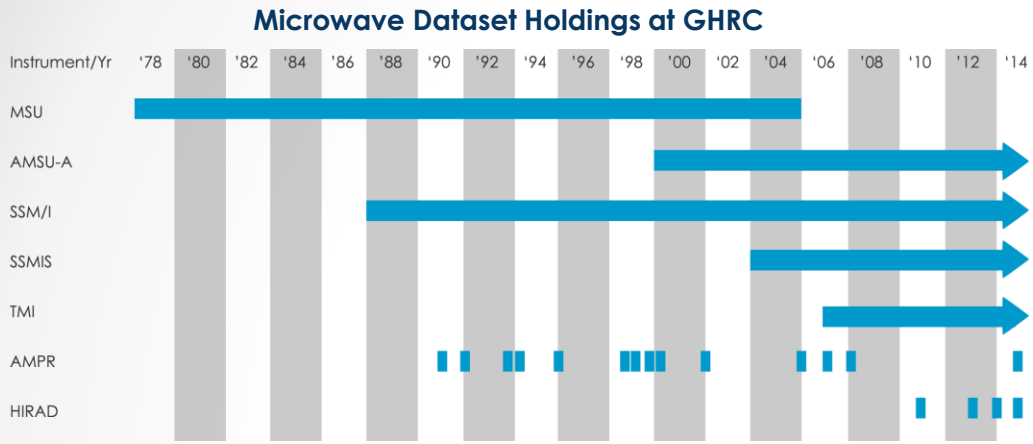
- Responsible for data from the TRMM Lightning Imaging Sensor, plus ancillary lightning data sets utilized by the LIS SCF scientists, since January 1998. A second LIS instrument will fly on the SpaceX rocket to the International Space Station in February 2016.
- Ancillary data –
  - National Lightning Detection Network, electric field mill data from the Kennedy Space Center, global infrared data and ground based radar data
- Precursor satellite instruments -
  - Optical Transient Detector in operation on Microlab-1 from 1995 to 2000
  - Operational Linescan Sensor on Defense Meteorological satellites from 1973 to 1995

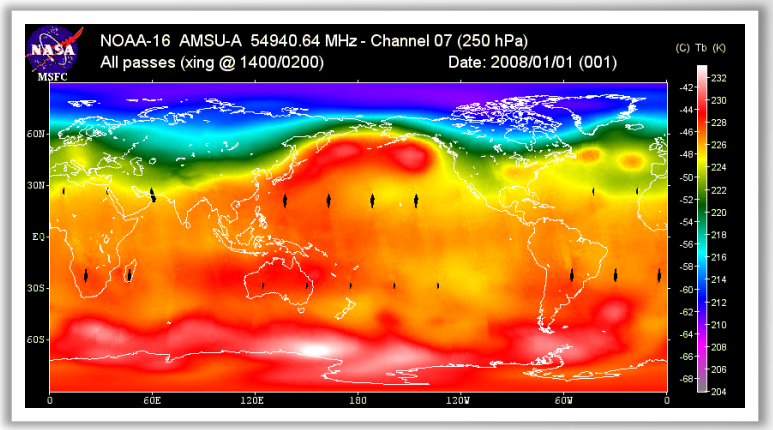


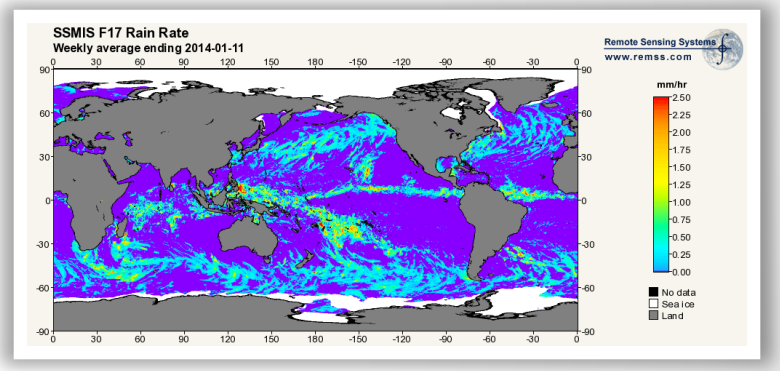


GHRC is recognized as the National Lightning Archive

What we serve
Microwave Data



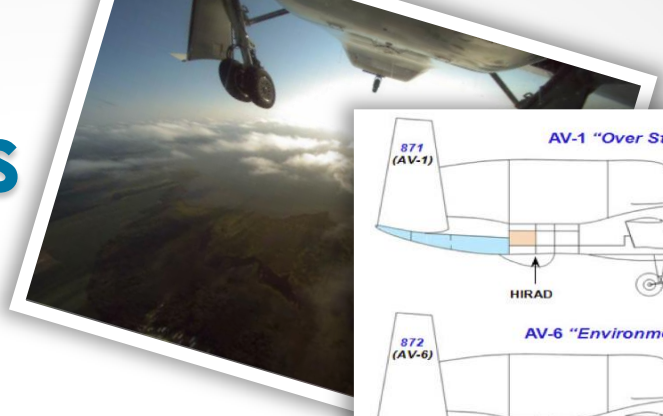


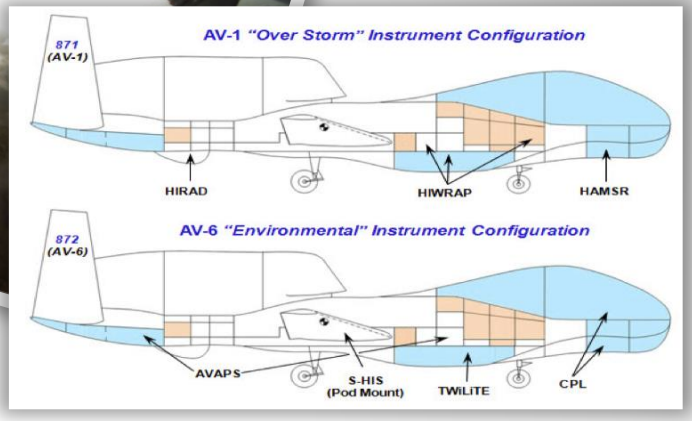


- GHRC and its predecessor programs have been ingesting, processing, archiving and distributing microwave data for over 35 years
  - MSU, SSMI, AMSU, AMPR, TMI, AMSR-E
- This climate sensitive data record extends back to 1978 providing an unbroken inventory of climate information that continues today

GHRC is also recognized as one of the primary data centers for microwave data

What we serve
Field Campaigns





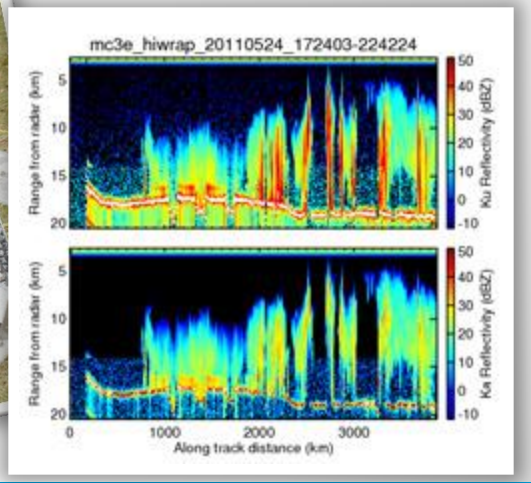
Hurricane Science

Data from successive field campaigns since 1990 are tied together through common procedures, consistent metadata, and discovery and archival systems making it easy to access data from instruments that have been employed across several missions

Hurricane and Severe Storm Sentinel (HS3)

Five-year mission to investigate the processes that underlie hurricane intensity change in the Atlantic Ocean basin and will utilize two Global Hawks

GHRC is recognized as one of the main data centers for Hurricane Science data



Global Precipitation Measurement Mission (GPM) Ground Validation (GV)

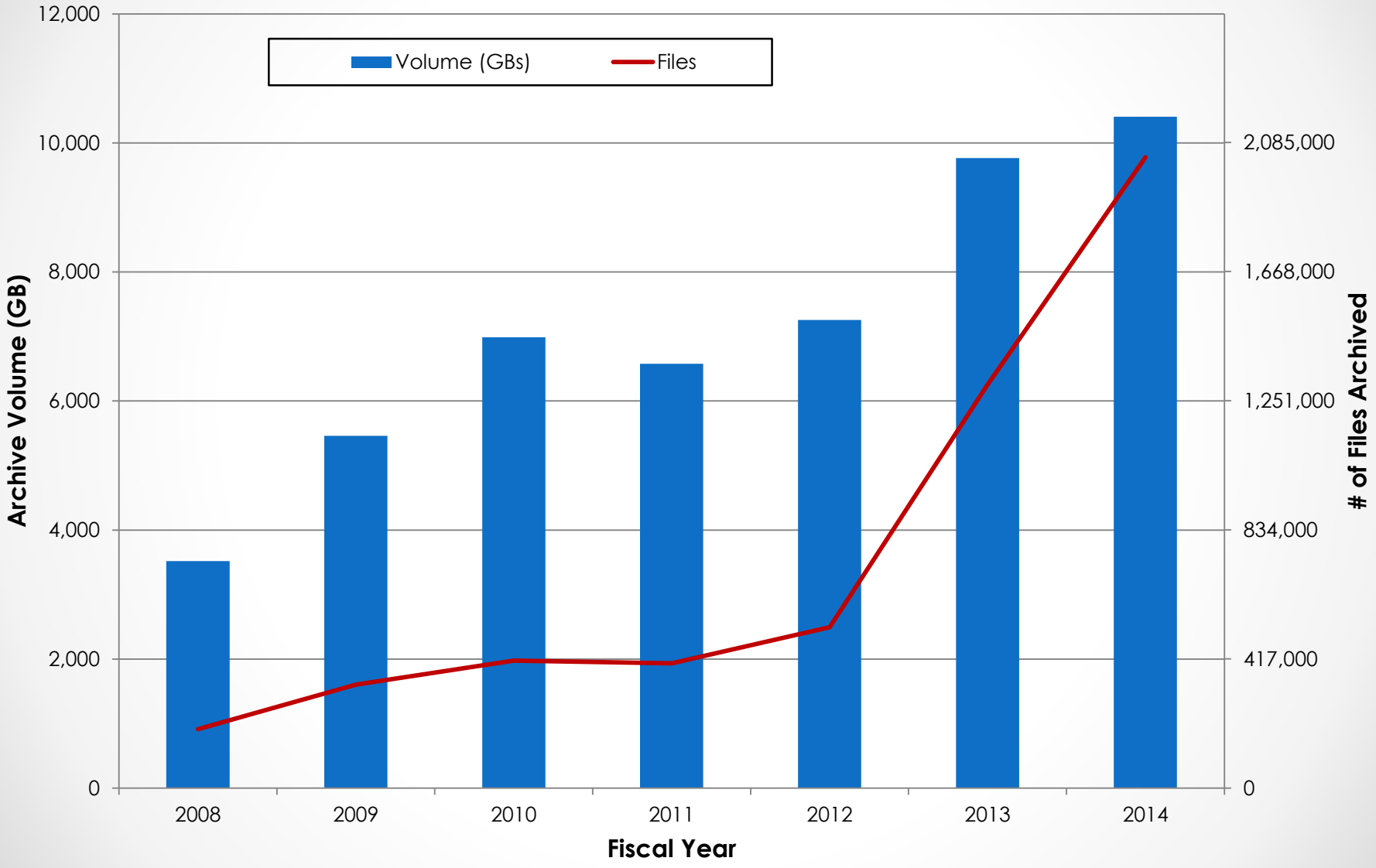
Ground and airborne precipitation datasets supporting physical validation of satellite-based precipitation retrieval algorithms

GHRC is set up to manage a large number of episodic, heterogeneous datasets and can handle the "long tail" of science data

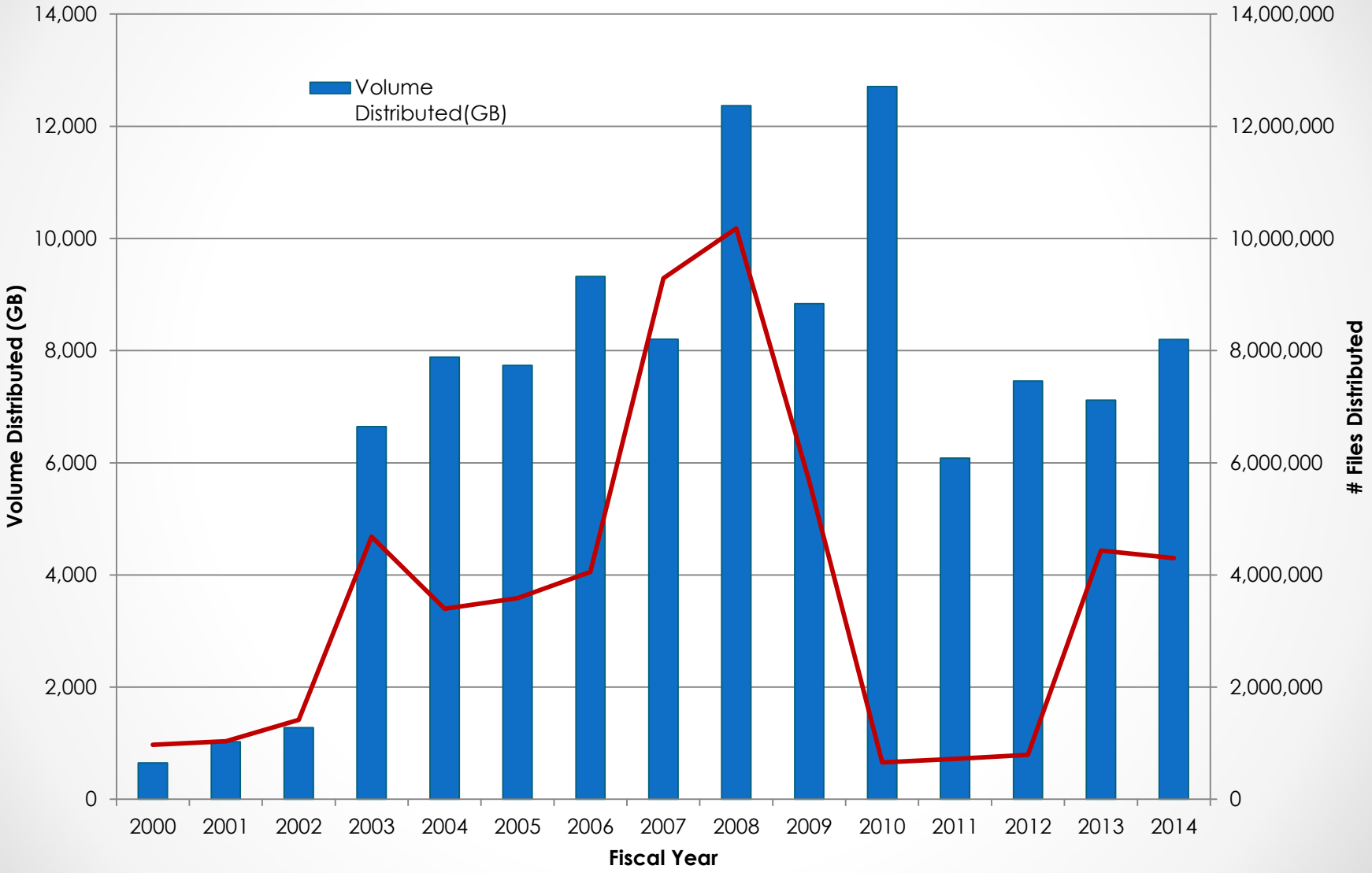
GHRC By the Numbers

- Registered Users – 1341
- Data sets
  - 291 Public
  - ~ 8% of total ESDIS holdings (3666)
  - 8 Limited visibility
  - 34 input data streams only used in processing to produce the final products
- Granules
  - ~ 2 Million (Archived since 94)
- Archive size - ~ 10 TB
  - HS3 will add 60 TB
- Distribution
  - ~82 million files since 94

GHRC Yearly Cumulative Archive (Oct 1, 2008 - Aug 29, 2014)



GHRC Yearly Data Distribution



Data Impact beyond Science Teams

Tellus
SERIES A DYNAMIC METEOROLOGY AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Tellus (2010), 62A, 393–410
Printed in Singapore. All rights reserved

TELLUS

Improving passive microwave sea ice concentration algorithms for coastal areas: applications to the Baltic Sea

By NINA MAAß* and LARS KALESCHKE, *Institute of Oceanography, University of*
20146 Hamburg, Germany

Science in China Ser. D Earth Sciences 2005 Vol.48 No.2 219—229 219

Climatological distribution of lightning density observed by satellites in China and its circumjacent regions

MA Ming, TAO Shanchang, ZHU Baoyou & LÜ Weitao

Science in China Series D: Earth Sciences
February 2005, Volume 48, Issue 2, pp 219-229

Climatological distribution of lightning density observed by satellites in China and its circumjacent regions

Ming Ma, Shanchang Tao, Baoyou Zhu, Weitao LÜ

Title: Preliminary Lightning Observations over Greece

Author: Themis G. Chronis

: Hellenic Center for Marine Research (HCMR), Institu

46.7 km Athens - Sounion Av., P. O 19013 Anavyssos

ELSEVIER

Atmospheric Research

Volume 91, Issues 2–4, February 2009, Pages 438–452

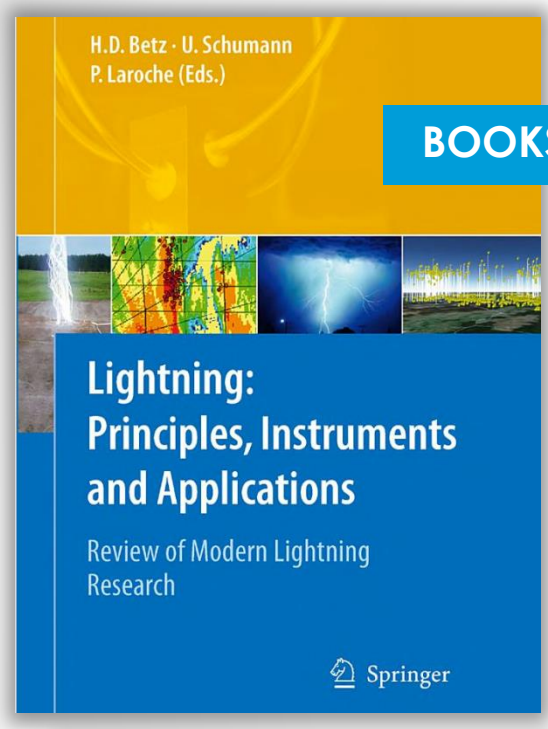
13th International Conference on Atmospheric Electricity — ICAE 2007

A comparison of lightning activity and convective indices over some monsoon-prone areas of China

Jianhua Dai[a, b], Yuan Wang[a], Lei Chen[b], Lan Tao[b], Jianfeng Gu[b], Jianchu Wang[c], Xiaodong Xu[c], Hong Lin[b], Yudan Gu[c]

GHRC data is used to address specific regional needs all over the world

Data Impact beyond Science Teams



BOOKS

Science
AAAS

Terrestrial Gamma-Ray Flashes Observed up to 20 MeV
David M. Smith *et al.*
Science **307**, 1085 (2005);
DOI: 10.1126/science.1107

New science areas

Science
AAAS

Direct Measurements of the Convective Recycling of the Upper Troposphere
Timothy H. Bertram *et al.*
Science **315**, 816 (2007);
DOI: 10.1126/science.1134548

WILDERNESS & ENVIRONMENTAL MEDICINE, **23,** 260–269 (2012)

WILDERNESS MEDICAL SOCIETY PRACTICE GUIDELINES

Wilderness Medical Society Practice Guidelines for the Prevention and Treatment of Lightning Injuries

Eric Johnson, MD; Scott E. McIntosh

VOLUME 11 JOURNAL OF HYDROMETEOROLOGY AUGUST 2010

The Hydrology and Hydrometeorology of Flooding in the Delaware River Basin

JAMES A. SMITH, MARY LYNN BAECK, AND GABRIELE VILLARINI

Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey

WITOLD F. KRAJEWSKI

IIHR–Hydroscience and Engineering, The University of Iowa, Iowa City, Iowa

(Manuscript received 13 November 2009, in final form 22 March 2010)

Science Applications

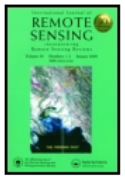
International Journal of Remote Sensing

Volume 31, Issue 13, 2010

Special Issue: Satellite Observations of the Wenchuan Earthquake of 12 May 2008

Select Language

Translator disclaimer



Precursory signals using satellite and ground data associated with the Wenchuan Earthquake of 12 May 2008

Ancillary data from GHRC is used in many, many papers

Our Vision for GHRC

- Efficient
  - Minimize any operational redundancies via automation
- Innovative
  - Design, develop and adopt new technologies to minimize cost and maximize productivity of our stakeholders
- Agile
  - Respond to changing needs (science driven/programmatic)
- Active
  - Collaborations with our stakeholders
  - Leadership roles in ES Informatics

http://inspirationalstorytellers.com/wp-content/uploads/2013/05/future-vision.jpg

Flat budgets

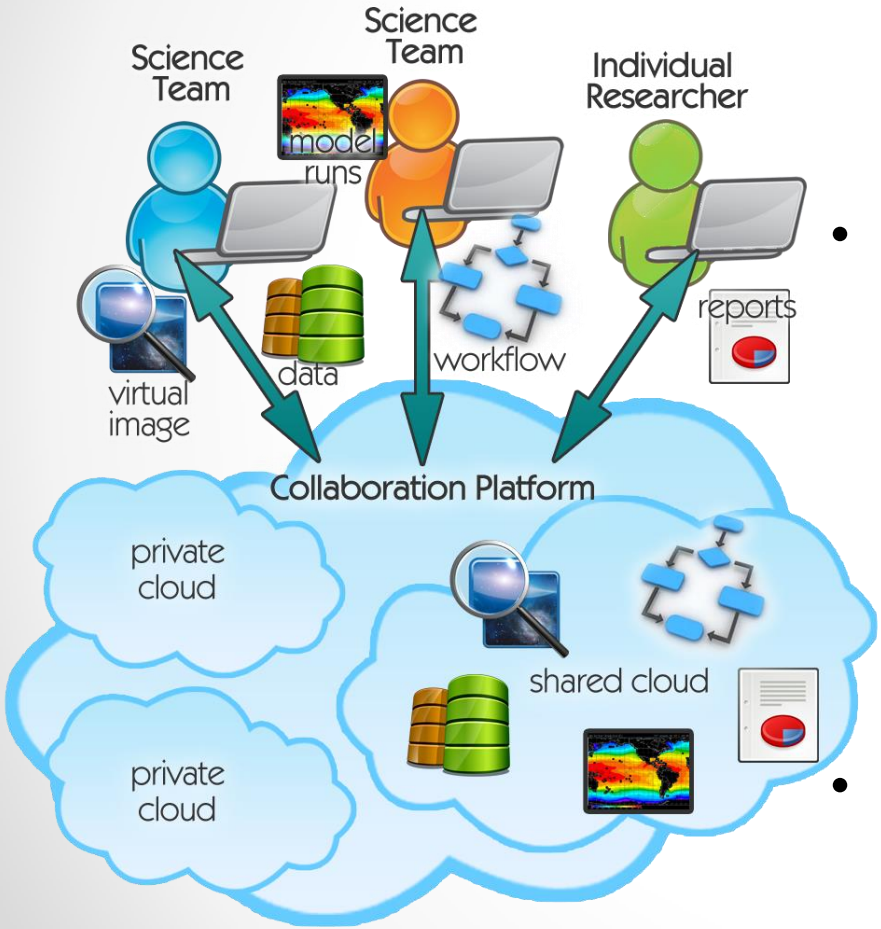
Community/Leadership Activities

- ESDSWG: focuses on community driven recommendation for Earth Science data system
  - Innovations Lab Working Group (McEniry, Ramachandran)
  - Airborne Working Group (Conover)
  - ASCII for Science Data (Conover)
- ESDIS Standards Office (Conover)
- IEEE GRSS Earth Science Informatics Technical Committee – (Ramachandran)
- AGU Earth and Space Science Informatics (ESSI)Focus Group

New Projects

- Computational Modeling Algorithms and Cyberinfrastructure (CMAC) program: *Collaborative Workbench to Accelerate Science Algorithm* Development (PI Ramachandran/UAH PI Maskey)
- Advanced Information Systems Technology (AIST) Program: *Automated Event Services: Efficient and Flexible Searching for Earth Sciences Phenomena* (PI Clune GSFC/Co-I Ramachandran)
- HS3 Data and Information System
- White House OSTP led Climate Data Initiative (CDI)
- White House OSTP Big Earth Data Initiative (BEDI)

Earth Science Collaboration Workbench (CWB)



- Augments a scientist's current research environment to allow him or her to easily share diverse data and algorithms
- Leverages technologies such as the cloud and social collaboration frameworks for scalable and controlled collaboration
  - Open source Eclipse framework, compatible with widely used scientific analysis tools such as IDL and Python.
- Misc.:
  - GLM Validation and Verification Tool
  - Provenance Service

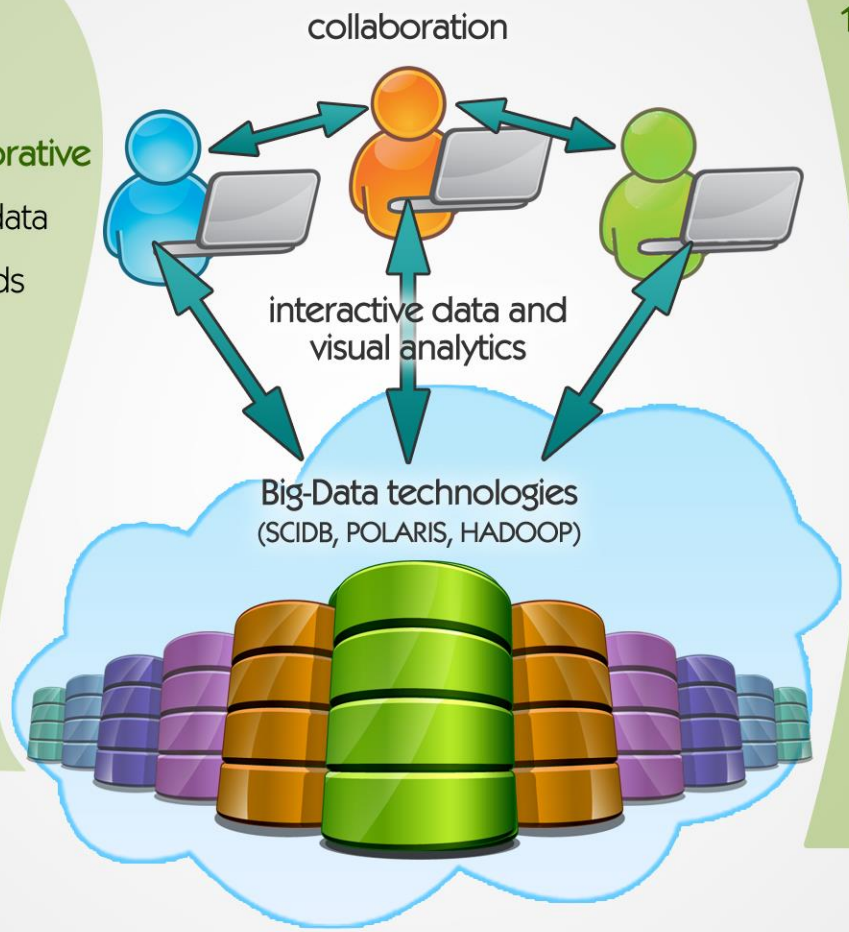
Automated Event Services

Utilize **Big-Data** technologies to…

- Enable **interactive** and **collaborative** scientific **data analysis** on big data
- Share data and analysis methods seamlessly,

…in order to…

- Relieve scientists from data management,
- Empower scientists to focus on science, and
- Boost science productivity.



1 **Identify occurrences** (events) of phenomena
   - Entities in the 4D spatiotemporal space

2 **Associate** additional relevant data with events.

3 **Characterize phenomena** with defining features extracted from data.

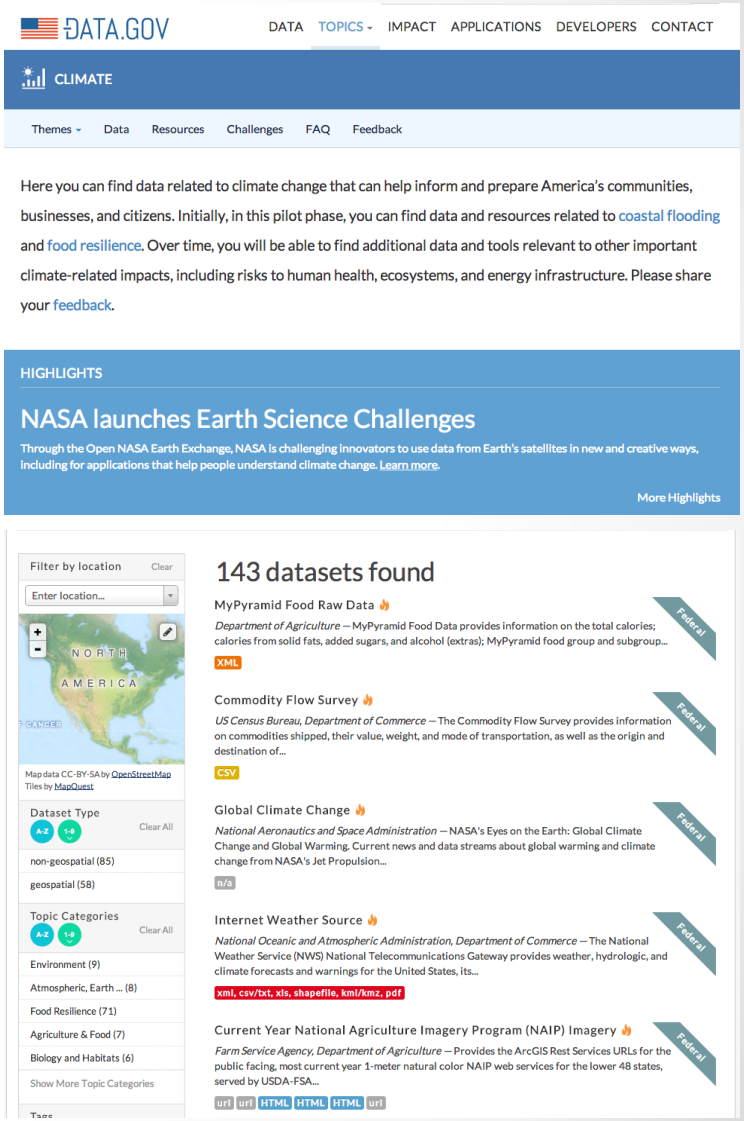
4 **Correlate** defining features of various phenomena in both space and time.

5 **Improve predictions** of future events using correla tions among phenomena for **better decision making**.

Big-Data Vision → **Technology Infrastructure** → **Science Enablement**

Climate Data Initiative

- NASA is leading the Climate Data Initiative being coordinated by the Council on Environmental Quality (CEQ) and the Office of Science and Technology Policy (OSTP).
  - Identify and make interoperable relevant data from multiple interagency sources to support climate
  - Facilitate the integration and better use of data for decision support and actionable science information
  - Make these data more accessible, discoverable, and usable for purposes other than which they were originally collected
- NASA has formed a Data Coordination team consisting of personnel from GHRC with appropriate expertise to support these goals.



Big Earth Data Initiative (BEDI)

- BEDI seeks to improve the discoverability, accessibility, and usability of data and information derived from Federal civil Earth observations, making these information products easier for everyone to find and use.
- *GHRC Task*
  - Data available online via services based on open standard protocols
    - Focus on Field Campaigns

Proposal Submitted

- Submitted to Advanced Information Systems Technology (AIST) Program
  - Developing a Numerical Weather Prediction and Data Dissemination Virtual Appliance to Support Disaster Preparedness, Mitigation, and Response (PI Molthan MSFC/Co-I Ramachandran)
  - DERECHOS: Data Environment for Rapid Exploration and Characterization of Organized Systems (PI Clune GSFC/Co-I Ramachandran)
  - GEODE: GEO Data Engine to Enable Big Data Analytics in Exascale-Computing Era (PI Ramachandran)
  - Illuminating the Darkness: Exploiting untapped data and information resources in Earth Science (PI Ramachandran)
- Pursuing other funding opportunities to build new capacity within GHRC

UWG Charge

http://www.marketingbrainfodder.com/files/2012/09/Inspire-300x199.jpg

- Select a scribe
- Elect a co-chair
  - Executive Session Friday Morning
- Provide prioritized recommendations/suggestions for improvements
- Tell us what we are doing right
- Write and submit a report

UWG Questions

- Data Stewardship
  - Are there additional important ancillary data that need to be in the GHRC catalog?
  - How can we make GHRC more visible to the science communities?
    - Science conferences/Meetings
- Knowledge Augmentation Services
  - Are there other services that GHRC can provide that will make your research process easier?
    - New means of data discovery
  - Should GHRC look at changing the access mechanisms?
    - Accessing data files from machine APIs (libs for python, idl)
    - Cloud based stores (AWS S3 for EC2 computation)
  - Need for new tools for data exploration and visualization?